

Analisis Pengaruh Seleksi Fitur *Information Gain* dan *Mutual Information* pada Klasifikasi Sentimen Ulasan Film Menggunakan *Support Vector Machine*

Hanif Salafi¹, Adiwijaya², Untari Novia Wisesty³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹hanifsalafi@student.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³untarinw@telkomuniversity.ac.id

Abstract

Sentiment analysis is an emotional identification technique that is expressed in the form of text. The purpose of sentiment analysis is to determine positive or negative opinions in a sentence or document. One way to determine the positive or negative opinions can be done by text classification. However, text classification has a problem, that is the number of attributes used in a dataset. Feature selection has an essential role in reducing the less relevant attributes of a dataset. Selection methods for features such as Information Gain (IG) and Mutual Information (MI) are methods commonly used in text classification. In this study, the domain used was a movie review. Each of the two feature selection methods is used as a feature selection to help improve the performance of the Support Vector Machine (SVM) classification algorithm. The results of SVM classification with the two feature selection methods indicate that the use of feature selection with SVM can increase the value of accuracy, where when not using feature selection only get an accuracy of 88.75%, while if added feature selection the accuracy increases to 89.05%. Then when the two feature selections are compared, Information Gain gives the same results as Mutual Information with the highest accuracy value of 89.05%. Thus, the use of Information Gain as a feature selection on the SVM classification algorithm gives the same results as Mutual Information.

Keywords: Sentiment Analysis, Classification, Movie Reviews, Information Gain, Mutual Information, Support Vector Machine
